DOCUMENT RESUME

ED 187 730 TM 800 225

AUTHOR	Airasian, Peter W.
TITLE	A Perspective on the Uses and Misues of Standardized Achievement Tests.
INSTITUTION	National Council on Measurement in Education, East Lansing, Mich.
PUB DATE	79
NOTE	13p.
AVAILABLE FROM	National Council on Measurement in Education, USPS 823120, 1230 17th Street, N.W., Washington, DC 20036 ($1.50).
JOURNAL CIT	NCME Measurement in Education; v10 n3 Fall 1979
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Accountability; *Achievement Tests; Decision Making; Educational Objectives; Educational Policy; *Educational Testing; Elementary Secondary Education; Measurement Techniques; Scores; *Standardized Tests; *Testing Problems; *Test Interpretation; Test Reliability; Test Results; Test Validity

ABSTRACT

The nature and common uses of standardized achievement tests are described and placed in context in order to evaluate claims and criticisms that are commonly made. Tests are described as samples of behavior and hence subject to sampling errors. Much of the debate is dominated by emotion and prejudices. There is very little empirical evidence about the effects of tests as a part of the complex educational situation. Test results are misused if they are treated as exact and permanent indices of achievement; if they are used as the sole basis for making important decisions; if the constructs they measure are misinterpreted; or if they are interpreted as caused only by the content of the test. Within the classroom, test scores may legitimately be used as supplements to other information available to the teacher. Outside the classroom, where other information may be less available, the use of tests for accountability decisions is more questionable; but the demands for student and teacher accountability are strongly supported and there may be no other, more credible, information available. The school and classroom are evaluative environments and evaluations will continue to be made. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
NCME
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED187730

TM800 225

Volume 10, No. 3 Fall, 1979

A Perspective On The Uses and Misuses of Standardized Achievement Tests[1]

Peter W. Airasian

About This Report

Peter Airasian

A recent sweep of the test interpretation literature covering the period 1922-80 has brought home to this editor that there truly is little new under the sun. While there have been significant advances in computer technology and in the types of score reports generated by this technology, the way we use and misuse tests blends from one decade into another. Much has been written in the pages of *ME* and in other journals concerning those issues which have captured the attention of test specialists; criterion vs. norm-referenced testing and reporting, objectives-based instruction, teachers' attitudes towards testing, the "dehumanizing" influences of testing and the like. But to one who has journeyed back into the literature of the past sixty years, these contemporary issues are really not contemporary at all. The axiom (however poorly paraphrased) that states, "One who is ignorant of the past is doomed to dwell in it" is most applicable to those of us in educational measurement who ignore what our professional forebearers learned long ago.

Peter Airasian *is* aware of the past and writes this paper with a clear and objective understanding that the uses and misuses of tests have long antecedents and do not appear to be as they are. His position will please few in the testing community for he scores both the proponents of tests and those who ascribe all types of horrors to them. But Dr. Airasian's objective analysis of test use and misuse cannot be ignored. He speaks with the clarity of one who has researched the topic here and in Ireland. He writes with objectivity and not out of a flair for polemics. The reader will surely think about this paper long after it has been finished and put away. Dr. Peter W. Airasian is a Professor, Division of Educational Research, Measurement and Evaluation, Boston College, Chestnut Hill, Massachusetts.

HCR

The present debate over the uses and misuses of standardized achievement tests does a disservice to the complexity of the issues involved. The rhetoric, and that is what it largely is, that produces claims and counterclaims about the uses and misuses of standardized tests has taken on a life of its own. Battle lines have been drawn and for many participants in the controversy there is no middle ground of uncertainty; if you're not for their position on testing you must be against it. In articulating their positions, both proponents and opponents of standardized testing have been guilty of distorting the historical development of testing and of failing to recognize that there is precious little empirical evidence to support claims of test use or misuse. The questions being raised about testing today are no different in origin from the questions that have been and will be raised about a variety

1. Revised version of an address prepared for a regional conference on testing and evaluation under a contract from the National Institute of Education. The views expressed are solely the responsibility of the author.

of other educational policies and practices, in that they are rarely raised at a time when it is reasonable to obtain empirical evidence about their answers. The furor over standardized testing in the United States did not arise until testing had so inundated education and society that it was difficult to identify groups which could help to clarify the effects, uses and misuses of such tests.

It is, however, one matter to distort or over-generalize historical and empirical data, but it is quite another to single out one variable, a standardized test, from the schooling context and to treat that variable as a prime motivator of educational practice. It may be simple to set up artificial dichotomies pitting teacher judgments against standardized test information in search of some "best" method. It may be simple to make generalizations about various test uses and misuses without considering the nature of and differences among classroom and school contexts. It also may be simple to make unqualified generalizations about the use and effects of tests on 'teachers' or 'principals' or 'students.' It may be simple to do these things, but it is also counterproductive, insofar as it distorts reality and prevents many from obtaining a full understanding of tests, testing, and their uses and misuses.

In preparing this discussion, a great deal of time was spent seeking a path through the jungle of commentary on the uses and misuses of standardized achievement tests. From this commentary efforts were made to identify a clear list of test practice and malpractice. However, attempts to understand the furor over standardized achievement tests on its own written merits were not productive. In order to make sense of the claims and counterclaims about standardized achievement tests, it was necessary to step back from the immediacy of many written sources and to consider the debate on more general grounds. For example, it was necessary to keep firmly in mind the definition of a test and the implications of that definition for the frequently voiced concerns over the effects of standardized achievement tests and the less frequently voiced, yet equally real, concern over the effects of so-called alternative testing procedures. It was also necessary to examine the context of schools and classrooms in considering the nature and extent of various claimed uses and misuses of tests. The author constantly found himself falling into a 'language trap' and in need of a continuous reminder that much of the language associated with tests and testing is emotional and value laden. Language not only reveals, it also conceals and in the latter respect makes interpretation sometimes tricky and oft-times itself value laden. In short, it was necessary to reflect and to try to understand the place of standardized achievement tests in education before it was possible to speak reasonably about the uses and misuses of such tests.

A primary aim of this discussion is to provide a perspective in which to place the current debate over standardized achievement test use. Examining the claims and counterclaims regarding test use in the context of this perspective makes it clear that there are problems in the positions advanced both by groups in favor of and opposed to testing. Certainly there are uses for standardized achievement tests and just as certainly there are misuses of these tests. However, in seeking to identify such uses and misuses two points need to be stressed. First, standardized achievement tests are neither as harmful nor as beneficial as commentators have painted them; the truth, as is so often the case, lies somewhere in the middle. Second, uses and misuses of tests are rarely absolute; they must be considered within the framework of other contextual variables. This presentation is neither a defense nor an indictment of tests; it is a plea for reflection and rational thought in the face of a debate that threatens to force us all into emotional rather than intellectual responses.

What is a Test?

It is appropriate to start by considering what an achievement test is. In the simplest sense an achievement test is a *sample of behavior that is used to make inferences about a larger domain of behavior.* This fairly short

definition contains substantial food for thought. First and foremost, a test is a "sample of behavior." No test, therefore, provides complete and exhaustive evidence about an individual's achievement or performance. Consider, for example, a test of pupils' ability to perform long division problems with remainders. This is just one of many skills which might be included in a fourth or fifth grade mathematics curriculum. There is literally an infinite number of long division problems with remainders which might be presented to pupils in order to test their learning. Obviously, no test constructor would — or could — construct a test containing every conceivable problem in order to measure pupils' learning of this skill. Instead, the test constructor would select some limited number of problems, a sample, and measure students' performance on this sample. On the basis of this sample of students' behavior, inferences — that is, generalizations — are made about how pupils would perform on the larger domain containing all possible long division with remainder problems. When we say that a student can solve long division problems with remainders, conjugate French verbs, capitalize and punctuate correctly, understand the Periodic Table, or write a coherent paragraph, we are basing such conclusions on observation of only a sample of the student's behavior. From this limited observation we make inferences regarding the student's general mastery of the behavior in question. From this definition of a test, three interrelated issues emerge.

Validity

First, the issue of test validity, with its various meanings and methods, turns out to be concerned essentially with the nature of the inferences a particular test permits. Thus, for example, when we examine the content validity of an achievement test by comparing the objectives of instruction to the items contained in the test, we are actually asking whether there is sufficient correspondence between what was taught and what is being tested to permit valid inferences about pupil learning. In the context of achievement testing, we rarely are interested in a specific set of items which comprise a test. Rather, we are concerned with these items as indicators of a larger, unsampled domain of items, since we wish to make inferences about the student's ability to perform on this larger domain, not simply on the particular items which comprise a given test.

Sampling Error

A second issue which derives from the preceding definition of a test concerns the topic of error. No test provides an exact inference about a pupil's performance on the larger behavior domain of interest. Because a test is a sample of behavior, in many cases an extremely small sample consisting of only one or two items for any given skill or objective, the results of a test are subject to error. If twenty people were to construct a 10-item test to assess pupils' ability to solve long division problems with remainders, it is unlikely that any two would construct identical tests. Further, if each of the twenty tests were to be successively administered to the pupil, it is probable that the pupil would not obtain an identical score on each. Some tests might contain a higher proportion of more difficult items than others. Distractions during test administration or careless computational mistakes might cause the student's performance to vary from one 10-item test to another. In most cases, we would not expect student performance to vary a great deal. However, we would probably not be surprised if the student's highest and lowest score over all twenty tests differed by two or three points. The nature of the particular sample of items which comprise a test as well as factors which occur during test administration (e.g., distractions, careless computational errors, etc.) can affect the stability of a pupil's test score.

Since most tests are administered only once to a pupil, we generally have no way of knowing to what extent such errors are reflected in the pupil's test score. Nor do we know for certain that a readministration of the same test or some equivalent form of the test would produce exactly the same score. It is because of a concern for such uncertainty that the issue of test reliability is important, insofar as reliability estimates provide an indication of the amount of error associated with test scores. By virtue of being a sample of behavior, only one of many possible samples, the results obtained from a test cannot be considered to be an exact, invariant index of pupil performance. No test provides this level of precision and all test results should be interpreted with this fact in mind.

The Test — A Broadened Concept

A third issue that derives from the definition of a test is that the word 'test' is a generic term, encompassing not only the standardized tests which are the primary focus of this discussion, but also numerous other educational situations in which a sample of behavior is used to make inferences about an individual or group's characteristics. There is a tendency in the literature on testing to confine the use of the word 'test' to commercially available norm-referenced standardized tests and to consider other evaluative situations as being different in kind. It is important, however, to make a distinction between what a test is and the structure a given type of test imposes on the testing situation.

Observations of pupil behavior which lead teachers to infer that pupil A is 'bright' or has a strong 'self concept' while pupil B is 'dull' or 'unmotivated' are tests in the generic sense of the word. Most proposed alternative evaluation or assessment procedures such as pupil interviews or examination of work samples are tests as well. While they are often different from traditional norm-referenced standardized achievement tests in form and content, as well as in the way the results are presented, they are tests nonetheless. It follows then that new, alternative and non-standardized forms of assessment should be subject to the same level of scrutiny and be expected to meet the same criteria that standardized achievement tests are expected to meet. When we think about tests, the uses and misuses of them, as well as the effects that tests have on schools, teachers, and most of all pupils, we should not confine attention solely to standardized achievement tests.

Tests then, come in many varieties and are designed to assess many dimensions of performance. An achievement test, for example, is intended to measure the extent to which pupils have learned a body of content or a set of educational objectives which they have been taught. A teacher-made achievement test may be prepared, administered and scored in a different manner than a standard-

ized achievement test, but both instruments are tests. So too are teacher observations, pupil interviews, classroom recitation and a host of other evaluative activities which are used to assess behavior and performance in classrooms. The conditions under which behavior is sampled may not be as controlled as when a standardized test is used and the actual measurements derived may not be as numerically or normatively laden as with standardized tests, but it is clear that inferences about pupils are made on the basis of sampled behaviors. Some have endeavored to make a distinction between standardized tests and other types of tests in terms of their effects on teachers and pupils. Standardized tests are claimed to have a variety of insidious and detrimental consequences while teacher-centered testing techniques are thought to be comparatively benign. In this regard, it is most important to bear in mind that whether one is talking about standardized tests or some teacher-made form of evidence gathering, one is talking about samples of behavior, subject to error, used to make inferences about pupils.

The Evaluative Context of Classrooms

A substantial amount of space has been devoted to considering the definition of a test and the implications of that definition. This has been done for two reasons. First, to indicate that the most important aspect of any test is the nature of the inferences it permits and how concerns about test validity and reliability flow directly from this fact; second, to portray the social context of schools and classrooms as testing situations with or without the introduction of standardized achievement tests. It is now necessary to consider this second point in more detail, since in order to consider the uses and misuses of standardized achievement tests one needs to have some perspective on the context into which such tests are introduced and used.

The Classroom

The school and classroom are evaluative environments in which testing and inference-making are both integral and essential aspects. Teachers observe, measure and make inferences about pupils — and vice versa. Moreover, the nature of the measurements and the attendant inferences which characterize schools and classrooms are wide-ranging. They encompass pupils' academic performance and potential, as well as their motivation, self concept, interest, attitudes, values and manners as well. While it has been noted already that standardized tests are different in some regards from the type of testing which characterizes day-to-day life in schools and classrooms, it is important to recognize that if there were no standardized tests — if such tests were eliminated — a great deal of testing and inference making would still go on in schools and classrooms.

And of course, it should. One of the primary roles of a teacher is to guide students' learning. One cannot accomplish this task without testing in one way or another. A teacher must determine where a student is with respect to the goals of instruction, what the student's potential is for succeeding with one instructional approach as opposed to another, how well the student has learned what has been taught, and what grade the student deserves for his or her performance. Good instruction is predicated on testing, on observing samples of pupils' behavior, on making inferences on the basis of these observations, and subsequently, on implementing some instructional treatment on the basis of the inference.

The school and the classroom, however, are more than academic learning environments; they are also social environments in which a multitude of teacher-student and student-student interactions take place every day. Suppose that teachers could not observe and make inferences about the pupils in their classes, that the classroom's informal testing and evaluation network were rendered inoperative. What might happen to the nature of social interaction? Social interactions are based upon and facilitated by the knowledge each participant possesses about the other. In a very short time, perhaps a week or so, most teachers know that when student A asks to leave the classroom he is malingering while if student B makes the same request there is a genuine need; or if student C asks a question it is to score what are euphemistically called "Brownie Points," while student D's asking the same question is indicative of genuine puzzlement. (Brophy and Good, 1974.)

Without the store of information about pupils that a teacher quickly accumulates by observing and making inferences about behavior, social communication in the classroom would be exceptionally difficult. Each new social interaction would have to be taken at face value, with each participant's motive, interest and need unknown to the other. The situation would be analogous to that faced by a substitute teacher entering a classroom without the store of information about pupils possessed by the regular classroom teacher. Communication is difficult, motives are suspect, requests to leave the room become moral dilemmas, and methods to quiet students acting out in class are hit and miss. Testing, usually by informal observational means is essential to the classroom and school, not only for educative purposes, but also to facilitate social interaction. If there were no standardized tests, classrooms would still be testing environments, for the simple reason that testing and inference making in some form is necessary for the classroom to function at all.

It is within this perspective that the uses and misuses of standardized tests, as well as the effects that such tests have on the educative process and its participants must be considered. In this light, it is clear that even in the absence of standardized achievement tests, instances of testing, evaluation, inter-pupil comparisons, and incorrect inferences about pupils will not disappear from schools and classrooms. They might be made on different bases, they might be more difficult to detect, their effects on pupils might be more covert, but they would exist nonetheless. It is simplistic to argue or to hold out promise that with the abolition of such tests will come classrooms not concerned with testing, classrooms devoid of inter-pupil comparisons, classrooms not characterized by a competitive ethic, classrooms in which pupils are not labelled or mislabelled and so on. All these things will occur in classrooms without standardized tests, because the classroom itself is a testing, evaluative environment.

Using The Test

These last comments lead directly into the topic of uses and misuses of standardized achievement tests. The discussion thus far has been largely descriptive. As one moves directly into the arena of test use and misuse, however, discussion and interpretation cannot help but become more personal, value-laden and subject to the intrinsic biases of the commentator. It is fair to say as a final preface to the remainder of this discussion, that as regards standardized achievement tests, one person's use is often another's misuse.

Let us start with two general parameters which characterize virtually the entire debate about uses and misuses of tests. First, the arguments in favor of and opposed to particular test uses tend to be simplistic, overgeneralized, and based upon very little empirical evidence. The rhetoric has reached such a height that individuals blithely, with little challenge or criticism, make statements such as "Standardized achievement tests increase educational opportunities for minority students" on the one hand, to "Tests don't murder every child, only 20 percent of them" on the other. The use of standardized tests, among all the variables that operate in the school context, has been singled out as the *cause* (and recognize that cause is an exceptionally strong word in educational parlance) of a host of beneficial and deleterious educational and personal outcomes. For example, the mass of teachers and school administrators are often portrayed as lambs, innocently and uncritically accepting standardized test results as the most valid piece of evidence available about a child and, on the basis of that evidence, wreaking havoc — and worse — on pupils, their school experiences and future lives. In reality, teachers are neither as naive nor as blind to limitations of standardized tests as many commentators who purport to speak for teachers would have us believe. Moreover, there is the implicit claim, made explicit by some, that if only standardized tests could be abolished or would disappear from the educational scene, fear, mistrust, lowered self concepts, labelling, favoritism and the rest would magically disappear from schools and classrooms. It is to be hoped that the prior discussion about the classroom as an evaluative environment would at least make one question whether it is reasonable to focus on standardized tests as the single cause of insecurity, labelling and the like, and whether these in fact would disappear from schools and classrooms with the removal of standardized tests.

We do not know what causes what in classrooms and schools because we generally are not in a position to single out for investigation any one variable from the complex of variables that characterize the schooling process. To what extent do standardized tests cause fear, labelling, and favoritism? We do not know. Few empirical studies carried out in the United States provide a clear answer to this question. We have only recently begun to obtain some idea of how teachers feel about and use standardized tests in their classrooms (Airasian, 1979; Brophy & Good, 1974; Goslin, 1967; Beck & Stetz, 1979; Stetz & Beck, 1978). Many studies of teachers' reported attitudes toward and uses of standardized tests were carried out at a time when the educational scene was comparatively devoid of the use of tests for accountability, curriculum evaluation, awarding high school diplomas or allocating remedial funds. Do the majority of American teachers think that tests provide accurate information about pupil achievement? How do teachers' own observations and judgments about pupils correspond to pupils' standardized test performance? What is the relative importance teachers accord standardized achievement test results as opposed to their own judgments obtained by daily observation of pupils in making educational decisions about pupils? We have scant information to answer these questions because we have been so caught up in rhetoric that we have failed to pose them in any systematic manner.

The answers to such questions are important and can do much to enlighten discussion about standardized test use and misuse. The current controversy over the uses and misuses of standardized tests often ignores, or makes strong assumptions, about the beliefs, attitudes and opinions teachers hold about such tests. Thus, for example, individuals and organizations calling for the abolition of standardized achievement tests because of their perceived misuses and the impact of these misuses on instruction and pupils, assume implicitly that tests and test results are sufficiently important or powerful in teachers' eyes to produce the host of detrimental consequences claimed. It seems reasonable to postulate that the ways in which standardized tests are used in classrooms and the consequent effect of these uses on pupils, the curriculum, and classroom climate are related to the way in which teachers perceive standardized tests. Teachers who perceive standardized tests and the constructs they measure to be irrelevant, biased, unstable or inaccurate, would be less likely to utilize test results in a practical way than teachers who perceive the test in a more favorable light. Additionally, even if it were shown that most teachers do consider standardized achievement tests and test results to be accurate, unbiased, and important, it would seem necessary to consider also the weight teachers accord to test information relative to other available evidence in making decisions about pupils, prior to suggesting a causal link between standardized tests and various pupil characteristics and classroom consequences. Few critics — or advocates for that matter — have called for obtaining information on such issues being content to rely upon unsubstantiated generalizations that portray teachers in a not particularly favorable light. Note that this argument does not claim that standardized tests and test results are *not* subject to misuse by some teachers in some schools. It simply raises three questions that have been overlooked in much of the present commentary: (1) What are teachers' attitudes and opinions about standardized achievement tests? (2) To what extent do teachers actually use standardized achievement tests in their classroom practices and decisions? and (3) If there were no standardized tests, would teachers find surrogate means to carry out the same practices and decisions? It would be useful to have evidence to answer these questions.

A second parameter which influences the debate over the uses and misuses of standardized achievement tests concerns the movement of education into a position of primacy as regards public policy formation. When the author was young, his parents gave him advice that stood him in fairly good stead until recent years. They said that in a debate, one should challenge the argument itself, not the person who was espousing the argument. It seems today that this advice is less valid. As education has moved to center stage in American social policy planning, whether by virtue of the Great Society programs aimed at various minority groups or the present-day teacher and pupil accountability schemes which are so prevalent, test

results have the power to affect the lives and livelihood of many constituent groups in society. In this policy framework, and given the impact of policy decisions that might be based on test results, it is fair to say that many proponents and opponents of standardized achievement tests have hidden agendas in their positions on testing; agendas that go beyond public statements about specific test uses and misuses. Testing companies obviously have a vested interest in the continued use of tests, and consequently tend to focus on the perceived benefits of these tests. Some groups fear accountability based on the test results and tend to focus their arguments against tests on a host of perceived negative consequences. Others have competing forms of tests or alternative measurement strategies to sell and thus have an interest in eliminating competition for their wares.

The intent in raising this general issue is not to impugn the motives and stances of all individuals and groups involved in the controversy over test use and misuse. Rather the aim is to point out that at a time when policy decisions based on standardized test results — or any other variable for that matter — have the potential to impact upon financial or job security, it may be difficult to separate the argument advanced from the individuals or groups advancing the argument.

All of this appears to put us in the realm of what has been called forensic social science (Rivlin, 1973); that is, a situation like a court trial in which each side of an issue puts forth *all* the evidence it can amass in support of its position, but none of the contrary evidence. However, the word "appears" is used for a very good reason. Rarely does either side of the testing controversy provide us with evidence. Instead we are confronted with simplification, overgeneralization, hidden and semi-hidden agendas, and no demonstrated inclination to examine the controversy in a more rational, understanding light. After examining the claims and counterclaims about tests and their use and misuse, contending with generalization, simplification, hidden agendas and so on, one's reaction is to wish for someone to say "I don't know" or "I'm not sure" so that meaningful discussion can begin. To a very great extent it is quite reasonable to conclude that the biggest abuse of standardized achievement tests is the level of discourse at which debate on the issue is carried out.

With this background, it is now necessary to turn attention to more practical considerations, in particular, to specific criticisms of standardized achievement tests and to a list of misuses which relate to these criticisms.

Criticisms of Standardized Tests

Criticisms of standardized achievement tests are wide ranging. In many respects, the tests are handy targets for criticism. They are *external* to the classroom; they are *commercial;* they are *standardized;* they are largely *norm-referenced;* and they are *summative* in nature. For many observers, just these test descriptors, quite independent of actual test use or misuse, conjure up pejorative images of the corporate-liberal state, routinization and the assembly line, impersonality, competition and loss of individuality. Moreover, tests often are imposed on the school and classroom in order to satisfy purposes external or peripheral to the perceived work of teachers and pupils, purposes which may impact upon and redirect the desired process of classroom education. Finally, standardized achievement tests do have limitations and are subject to misuse. For all of these reasons, tests are the subject of considerable criticism.

Any cursory literature review will reveal that there are dozens and dozens of criticisms of standardized achievement tests. This discussion cannot examine in detail every criticism voiced. Instead, standardized achievement test criticisms will be presented in terms of three general areas. Specifically, standardized achievement tests are criticized on the basis of characteristics intrinsic to the tests; that is, because of what they are or what they measure. The tests are also criticized on the basis of characteristics not intrinsic to them; that is, because of what they aren't or don't measure. Lastly, they are criticized on the basis of what they do; that is, because of their perceived effects on pupils, teachers, and the educative process in general.

Examples of criticisms related to characteristics *intrinsic* to the tests, include the following:

- — standardized achievement tests take for granted that everybody places equal value on the skills being tested.
- — standardized achievement tests assume a homogeneous culture and/or a common set of curriculum objectives.
- — standardized achievement tests are biased in favor of white, middle class children.
- — standardized achievement tests are constructed on the assumption that about half the pupils tested will be below average.
- — standardized achievement tests are impersonal, ambiguous, imprecise, undemocratic, and irrelevant.
- — standardized achievement tests measure only a small portion of what teachers and schools try to do.
- — standardized achievement tests are not valid measures of school-specific achievement.

Examples of criticisms related to characteristics *not intrinsic* to the tests, include the following:

- — standardized achievement tests are not diagnostic.
- — standardized achievement tests do not measure the goals of new, innovative, or "different" instructional interventions
- — standardized achievement tests do not measure creativity, interest, initiative, or values.
- — standardized achievement tests do not assess an individual's learning; rather they compare students' achievement.

Examples of criticisms related to the *effects* of the tests include the following:

- — standardized achievement tests "label" pupils.
- — standardized achievement tests "mislabel" pupils.
- — standardized achievement tests dominate school curricula and foster teaching for the test.
- — standardized achievement tests foster competition in the classroom.

— standardized achievement tests arouse fear and satisfy greed.

Within these three general domains fit the rationale for most claimed misuses of standardized achievement tests. For example, those who argue that it is a test misuse to evaluate teachers' performance on the basis of their pupils' standardized achievement test performance would — and have — sought justification for this position in all three domains of criticism. Thus, one version of the argument might proceed as follows: standardized achievement tests measure only a small portion of what teachers try to do (intrinsic characteristic), standardized achievement tests do not measure specific classroom goals (extrinsic characteristic), and, to use standardized achievement tests in this manner will mean that test content will dominate the classroom curriculum (effect). Perceived misuses of standardized achievement tests generally are portrayed in terms of one or more of these three domains of criticism.

Before listing major misuses of standardized achievement tests in light of these domains, it is important to make some general comments about the specific criticisms which comprise the domains. First, with a list before one, anomolies among specific criticisms appear. Thus, standardized tests are criticized by some because they measure an exceptionally limited and trivial dimension of the entire pupil. The same tests are criticized by others because test performance is based on too broad a range of pupil characteristics, including cultural adaptation, social class, vocabulary, standard English, language proficiency, the ability to work and think quickly, verbal skills, creativity, emotional and social security, and ability to function in a competitive situation. Second, it appears that many, though not all, of the criticism concerning the characteristics which tests do or do not possess, may be overcome in practice by the judicious and informed selection of tests in the first place. Third, given that any test is a sample of behavior, it should be apparent that no single test can capture the full richness of the human being. Indicting standardized achievement tests because they don't measure the whole individual is a little like the bait and switch tactic. Any test can be criticized for what it does *not* do, and this is something like criticizing the family car because it does not wash clothes. The logical response to this argument is not to throw the car away, but to buy a clothes washer. Similarly, one could reasonably take criticisms of what standardized achievement tests do not measure to be an argument for using more tests to measure additional pupil characteristics, rather than as an argument for abandoning standardized achievement tests altogether. Moreover, and perhaps most importantly, one may ask whether it is necessarily a bad thing that we do not explicitly measure each pupil's interest, initiative, values and self-concept, to name just a few. These traits may be considered private matters which should not be subject to scrutiny, particularly since we know very little about whether we can or should try to manipulate these in the educational setting. Explicit numerical information about the many pupil characteristics that critics claim are not measured by standardized achievement tests may be potentially more dangerous for both pupils and teachers than having information available only about more narrow, cognitive dimensions. That is an issue the reader may wish to consider sometime. What we may say in summary at this point, however, is that a criticism of a test is not isomorphic to a test misuse.

Misuses of Standardized Achievement Tests

In light of these considerations, what are the principal misuses of standardized achievement tests? Four such misuses will be considered. First, regardless of the context, type of decision being made, or individual making the decision, it is a misuse of standardized achievement tests to treat scores from these tests as if they were an infallible, unchanging index containing no error.

A test score, of whatever form or from whatever type of test, contains error, insofar as it is impossible to say with certainty that the individual's observed score indicates his or her "true" performance on the general domain to which we wish to make inferences. Even the most reliable tests available have standard errors of measurement equal to about 3.0, meaning that on the basis of a given test administration we can say with 95 percent probability that an individual's true score lies somewhere within a 12 point spread. Note the implications of this type of score interpretation as compared to interpretations which place absolute credence in one particular obtained score. It is this latter type of interpretation which leads to making different instructional decisions about a pupil who scores 112 and a pupil who scores 115, even though the test is not precise enough to indicate that the three point score difference represents a true difference between the individuals. We would be reluctant to infer on the basis of a single string of bowling that an individual who bowled 115 would, on all occasions, score higher than an individual who bowled 112. There is no reason why we should expect a single test administration to make sharp distinctions between pupils whose measured achievement is also within a few points.

A second misuse of tests, which is a corollary to the prior misuse, occurs when a single test score is used as the sole criterion for making important decisions about individuals and programs. Test results should always be considered and interpreted in light of various contextual parameters which characterize the pupils, teachers and general school environment. The score obtained from any test represents an abstraction, that is, a narrowing of the environmental factors which might influence that score. Thus, while it may be true that mean test scores of two classes indicate that one class performed substantially higher than the other, decisions based on such information should also take into account similarities and dissimilarities between the classes on other variables. The same is true when decisions are made about individuals, be these pupils or teachers. The preceeding caution is not intended to suggest that test scores are necessarily biased or inaccurate, only that they represent abstractions of what may be very complex situations or contexts. So too, for that matter, do most forms of tests which have been discussed thusfar. When important decisions which have the power to affect substantially the lives of individuals are to be made — be these decisions concerned with promotion, receipt of a high school diploma, merit raises, hiring or firing — they should not be made solely on the basis of a single test score.

In our everyday practical, political and bureaucratic world, there will always be a strong temptation and pressure to use performance on a single measure,

particularly a standardized test, as the basis on which various rewards are allocated or sanctions applied. The use of a cut-off score on a single index, especially one that is external to the classroom, school or school district, has appeal because it appears more "objective" to many observers and because it seems to remove "subjectivity" or political pressure from the decision-making process. In some sense these advantages may be realized, but they will be so at the expense of many pupils, teachers and schools whose unique circumstances will be ignored by a single test score. Including information in addition to a test score will not overcome all errors of judgments which might be made when a test score alone is used, but it will help reduce the number of such errors.

A third major misuse of standardized achievement tests occurs when uncritical assumptions are made about the traits that a test is presumed to measure. Just because a test has the word "achievement" in its title does not mean that for all pupils in all situations the test taps only or even exclusively achievement. For some groups, depending upon the nature of the instruction they have received, their verbal and English fluency, and the richness of their out of school experiences, any test, and a standardized achievement test in particular, may not be a measure of school achievement at all, but rather a surrogate measure of verbal fluency, home background or social status (Madaus, Airasian, and Kellaghan, 1980). At issue here is whether a standardized achievement test permits valid inferences about *school* learning. If students have not received instruction in the areas tested, if they are not fluent in English, or if they have reading difficulties, a so-called achievement test may not measure the pupils' school-specific achievement. If a pupil is confronted with situations he cannot understand or has not been exposed to in the course of school instruction, that pupil must rely upon extra-school experiences and knowledge in formulating an answer. Such extra-school experiences tend to be related to home background and social status. As noted previously, many of these problems can be overcome by judicious selection of tests in the first place. However, the point to bear in mind is that because a test is claimed to be an achievement test by the author or publisher does not mean that it is an achievement test in all circumstances and with all pupils.

A fourth misuse of standardized achievement tests occurs when the information provided by a test score is confused with interpretations of what caused the behavior summarized by that score. A test score is a numerical description; it is not good or bad in its own right, but only when it is interpreted in terms of some judgmental criterion. Nor is it a measure of the worth of an individual. A test score provides, albeit in a somewhat cryptic manner, an estimate of an individual's level of performance at a given point in time. A test score does *not* tell us why the individual performed as he/she did. The problems of making important decisions solely on the basis of a single test score was alluded to previously. This practice was labelled a misuse of testing because a single test score represents an abstraction from the contextual parameters which might relate to or influence it. For the same reason, it is extremely hazardous to try to explain the reasons underlying test performance without reference to various contextual parameters.

Many people in America perceive a test score decline in recent years. A variety of causal factors are cited to explain the perceived decline: scores are down *because* teachers aren't teaching the right things; score are down *because* students are lazy; scores are down *because* schools expect less of students nowadays. In many states pupils may be denied a high school diploma if their performance on a test is below a specified score (Airasian, Madaus, and Padulla, 1979). Implicitly, educational authorities in these states are saying that the reason a diploma will be denied to some pupils is because they did not learn what they were supposed to learn. The blame and explanations for poor performance is student specific. The causes underlying poor test performance are being interpreted and acted upon in a way which places the onus for failure squarely on the pupil's shoulders. Explanations such as faulty teaching, inadequate remedial programs, poor school resources, administrator's inability or unwillingness to reorient curricula in line with tested areas — all plausible explanations for the failure of some students — are being treated as though they were nonexistent.

Standardized achievement tests measure status, the performance of an individual at a given point in time. They do not indicate why the individual performed as he did. Moreover, no matter how expertly we combine and manipulate standardized test data with other data about schools and children, we can never reach anything more than probabilistic inferences about what caused what or about what makes one test score higher than another. Teachers must recognize this fact about both standardized tests and other forms of classroom measurement. Administrators, legislators and the public at large must do the same.

The nature of the educational process requires us to make decisions about pupils. We cannot ignore this fact. The four test misuses discussed are concerned primarily with situations in which educational decisions are made solely on the basis of standardized achievement test results. No single test, given at any single time and divorced from consideration of contextual contingencies should serve as the sole criterion for making decisions or providing explanations for pupil behavior.

Admittedly, these four misuses are general and do not attend to the many more specific claimed misuses of standardized achievement tests. There are three reasons for this generality. First, on reflection, it is likely that the reader will be able to place more specific claimed misuses within one of the four general areas discussed. For example, the issue of test bias, although not discussed specifically is related to misuses arising from the use of a test score as the sole criterion for making important decisions about pupils, as well as to misuses from making assumptions about what a so-called "achievement test" really measures and misuses from attributing causal explanations to observed test performance. Second, there are so many specific test misuses cited in the literature that it is impossible to consider each individually here. Third, and perhaps most importantly, it is not at all obvious that some claimed misuses of tests are in fact real misuses. Consider one currently popular example.

One issue specifically omitted from the four test misuses concerns the influence of test results in labelling children. Let us recognize that the word "labelling" has taken on a pejorative meaning and thus has an emotive connotation in the first place. Let us recognize that critics of testing always cast "labelling" in a negative, detrimental light while proponents of testing always cast it in a positive, enhancing light. Let us recognize that the empirical evidence about the influence of test information in the process of labelling pupils is equivocal at best. Let us

recognize that teachers label pupils in the absence of test information — they have to in order to facilitate academic and social interaction in the classroom. Let us recognize that the bulk of the research on the influence of test information on teachers' perceptions, Rosenthal and Jacobson (1968) in particular, is concerned with raised expectations and subsequent performance, while much of the present debate focuses on lowered expectations and performance. Most of all, let us recognize that labelling, based upon any type of information, is not intrinsically bad, incorrect or harmful. There is substantial evidence available which indicates that teacher ratings of pupils and pupils' test performance agree quite closely (Airasian, Kellaghan, Madaus & Pedulla, 1977; Brophy and Good, 1974). There is evidence, albeit more tentative, which indicates that when test information and the teacher's perceptions diverge for a pupil, teachers tend to give the benefit of the doubt to that piece of information most favorable to the pupil.

However evocative the terminology, the issue of the influence of test information on the labels attached to pupils is not at all clear. Standardized test results probably help and probably hurt some pupils. Whether the number of pupils who are helped or hurt as a direct result of test information is large or small is not known, nor do we know much about the impact of test information relative to other available evidence on pupils, such as teacher perceptions. Certainly test scores are related to the labels attached to some pupils, but the number of pupils at issue, whether or not the labels are beneficial or injurious to them, and the complex interaction between test results and teacher judgments remain to be explicated before we can clearly indicate that standardized achievement tests are powerful determinants of the labels attached to pupils. In some cases, it might be argued that for some pupils, namely those who are judged more favorably after testing, it is a misuse *not* to test.

Uses of Standardized Achievement Tests

Thus far, much attention has been devoted to considering test misuses. It is appropriate now to consider the topic of test uses. Like the discussion of standardized achievement test misuses, discussion of the uses of standardized achievement tests tends to be simplified, overgeneralized and not empirically based. If we paid heed to the most strident proponents of standardized test use, it would be impossible for the educational process to be carried on without such tests, so diverse are their claimed uses.

For the most part, however, we may place claimed standardized achievement test uses into two general categories: those concerned with the use of tests within classrooms and those concerned with the use of tests for extra-classroom purposes. Into the first category fit claimed test uses such as

- to provide diagnostic information about a pupil's or a class's learning problems,
- to assess the learning progress of pupils and classes,
- to provide information about pupils which can help teachers select the instructional approach most suitable for them,
- to provide data helpful in guiding pupils and their parents regarding various educational options; to predict future performance,
- to evaluate the overall effectiveness of the classroom program,
- to facilitate decisions involving grouping or classifying students, including "special needs" students.

By and large, these and other similar claimed uses relate to types of decisions which might be made primarily by teachers in the classroom setting.

Claimed test uses which fit into the second category, extra classroom uses, encompass the many forms of accountability which are seen with increasing frequency in American education. These are administrative or legislative uses of tests to compare schools and classrooms, allocate funds, award diplomas, evaluate teachers, compare program effectiveness, and alter curriculum. In many respects, it is the uncritical use of standardized achievement tests for accountability decisions which represents the most significant and at the same time perplexing development in testing in the past 20 years. This topic will be addressed shortly, but first some comments about the claimed within-classroom uses of standardized achievement tests are necessary.

The information one receives from a standardized achievement test, particularly one of the norm-referenced variety, tells very little about what pupils can or cannot do. The tests tap fairly general constructs with a relatively small number of items. They are not particularly useful therefore in providing precise diagnostic information about individual pupils, classes, or the success of the classroom instructional program, and certainly, in light of prior comments regarding test use, they should not be the sole basis for grouping students, or determining what instructional program a pupil will follow. They probably should not even be a part of any evidence used to make within-classroom decisions about pupils unless the content tapped by the achievement test used reflects the content of instruction.

Within the classroom, there is one primary use of standardized achievement tests. They provide an additional source of evidence about pupils which teachers may use to corroborate their own judgments and preceptions about pupils. Tests certainly are not always accurate in their assessments of pupils, but neither are teachers. More often than not, test results will tend to agree with teacher judgments, but in those cases where they do not, a red flag may be raised to suggest further consideration of particular pupils. Despite the fact that some commentators would wish us to believe that standardized tests and teacher judgments cannot and should not co-exist in the classroom, the two sources of evidence should not be considered as mutually exclusive or competitive. They each represent imperfect assessments of pupil characteristics and as such, when used in combination and subject to the misuses noted previously, they should better serve both the teacher and the pupil. Note that in some circumstances it may be a misuse not to rely upon tests or some other independent assessment to corrobroate teacher judgments.

The reader may correctly conclude from this short list that the uses of standardized achievement tests as the sole or even primary arbiter of decisions in the classroom are comparatively few. That does not mean, however, that the

corroborative, independent view of pupil achievement that the tests provide is unimportant. Left to their own devices, the vast majority of classroom teachers reach an accommodation with standardized achievement tests which puts the information from such tests in its proper perspective. The vast majority of classroom teachers the author has polled treat standardized achievement test information not as the be-all and end-all indicator of what a pupil really is, but as a piece of evidence that is sometimes helpful but more often not in carrying out their day to day activities.

The real problems with standardized achievement tests occur when they prevent, hamper or intrude heavily on the normal classroom routine, and in considering such intrusions, we are brought face to face with the extra-classroom uses of standardized test results for various types of accountability decisions. In this regard, the following generalization is appropriate: the stronger and more influential the intended use of test results, the stronger and more influential the tests become in affected persons' eyes (Bloom, 1969). Put another way, the greater the use of standardized achievement tests for making accountability decisions that transcend the classroom, the more the tests will intrude on classroom practices and instruction. In European countries, for example, the content of important certifying examinations such as the 'O' and 'A' level examinations in England, literally dictate the content and nature of classroom instruction (Madaus and Macnamara, 1970; Srinivasan, 1971).

When standardized achievement test results are used as the basis for staff reduction, merit raises, inter-school comparisons, allocating funds, assessing school effectiveness and awarding high school diplomas — as they are with increasing frequency in the United States — the tests and more particularly, pupils' performance on the tests, take on heightened importance. Since the decisions made on the basis of the tests have the potential to affect the life chances of pupils, teachers, and administrators by virtue of the sanctions imposed for poor performance, it is little wonder that a wide variety of responses to accountability occur, from teaching for the test to calls for the abolition of the tests themselves.

Rhetoric aside, the primary impetus for calls for test reform or abolition derives from the extra-classroom uses of tests in various accountability decisions, not from problems associated with strictly within classroom uses of tests. By making the results of standardized achievement tests important for various accountability decisions one makes the tests themselves important. By making the tests important, one encourages a variety of responses and alterations in practice to conform to the nature and content of the tests.

It would be easy, and certainly consistent with the list of misuses presented earlier, to say simply that any use of a standardized achievement test to make accountability decisions about teachers, pupils, administrators, or schools is inappropriate. But life is not quite that simple. While recognizing the many limitations in the use of standardized achievement tests as the primary basis for making accountability decisions, it is necessary also to recognize that such tests are perceived by the majority of the American public as providing the most credible, "objective" evidence about schools and learning. Yet standardized tests have serious limitations when used in accountability decisions and do not necessarily represent the appropriate basis on which to make such decisions (Madaus, Airasian, and Kellaghan, 1980).

However, like it or not, we live in a normative world, a world in which comparisons among individuals and institutions are not only wanted, but actively sought. Taxpayers, parents, administrators and legislators — for a variety of purposes — wish to make comparisons between pupils, teachers and schools. For the most part, these publics are unable or unwilling to accept testimonials about the goodness of education from those whom they perceive to have a vested interest in that goodness, namely teachers and administrators. These publics want some external, objective assurances of the quality of learning and teaching, and so they turn, virtually by default, to the one index they believe provides the external objective evidence they seek; standardized achievement tests.

The public wants and has a right to ask questions about the effectiveness and adequacy of its social institutions. The public has a right to seek assurances that its schools are doing a satisfactory job. And given the sanctity of local control which characterizes American education — and which makes American education unique in the world — there are no national syllabi, no common goals, and no common tests which serve as convenient methods for making the desired comparisons. And so we pay a price when we ask accountability questions; that price being the use of standardized achievement tests to perform a task they were not designed to perform and hence do not perform very well. But the tests are the only indices available which have credibility with much of the public, have a political validity with legislators and administrators and are developed by agencies external to the school. Standardized achievement tests for these reason represent a "politically" attractive index on which to base accountability decisions.

Thus far nothing explicit has been said about the use or misuse of standardized achievement tests when applied to extra-classroom accountability decisions. The reason for the silence on this issue is because the author finds himself on the horns of a dilemma. On the one hand it is a misuse to employ standardized achievement tests as the sole criterion for making decisions about promotion, tenure, school effectiveness, and high school graduation, since inherent in the use of these tests for such decisions are a variety of problems alluded to throughout this paper. On the other hand, the public has a right to know, measure and evaluate individuals and aspects of its schools. Moreover, at present the most politically viable, if not the most technically appropriate, index for such public disclosure is standardized achievement test results. In many locales, standardized achievement tests are used for accountability decisions virtually by default, because there are few viable alternatives. Accountability simply is not going to go away; if anything it is going to heighten and broaden in the next few years. For many reasons, we do not have the types of instruments available today which will enable us to satisfy in a fair and rigorous manner accountability demands. What is needed is a simultaneous clarification of school goals and an examination of old and new forms of tests in terms of their suitability to assess these goals.

One response to this duality is to categorize the reliance upon standardized achievement tests for extra-classroom accountability decisions as both a use and a misuse. However, it is hard to be happy with such a categorization, either philosophically or practically. It is clear that the extra-classroom uses of standardized achievement tests lie at the heart of the debate over testing. People who are

likely to be affected adversely by such a use have a vested interest in curtailing that use. It is also clear that the change from using standardized test results to aid the teacher in the classroom context to using such tests in broader policy oriented, extra-classroom decision making, is the most significant development in testing in the last two decades. Moreover, the movement toward accountability has heightened the importance of achievement tests in American education and as a result, will increase the tests' influence on classroom instruction and learning. We know why standardized achievement tests are used for these purposes, but recognize their limited usefulness.

One recent and promising alternative to the selection of standardized achievement tests for accountability decisions virtually by default, is occuring in many school districts and states in response to the minimal competency testing movement. Groups of educators and citizens are working together to define educational priorities for the schools and to devise appropriate teaching and testing strategies in light of these priorities. Such an approach overcomes many of the problems noted above because the issue of testing for accountability is being considered as an integral aspect of goal definition and instruction, not as an "add-on" or disjoint feature. When choices about the nature of the tests to be used for accountability are recognized as being in large measure dependent on the goals and instructional process for which individual schools are being held accountable, much of the furor over accountability will lessen and the accountability systems themselves will become more humane. It is not necessarily bad that tests influence what goes on in classrooms if the tests reflect agreed upon and achievable goals for classrooms. If there is a middle ground in the debate over the benefits and liabilities of extra-classroom accountability, it likely will be found in the understanding that goals, instruction and testing are integrally related and therefore must all be considered in framing accountability schemes. Perhaps all that can be done in this general discussion is to hope that with wider understanding and awareness of this issue, decisions about whether, and under what conditions employing standardized achievement tests for extra-classroom accountability decisions represent a test use or a test misuse will be better informed.

Conclusion

There is much left to say about standardized achievement test use and misuse, not only in elaboration of points addressed thus far, but also in terms of more specific topics which have not been discussed (e.g.; criterion versus norm-referenced testing). This discussion has tried to provide a glimpse of the breadth of issues associated with test use and misuse, but in so doing has sacrificed depth to breadth.

To conclude, the following six points are important. A test, any test, is a sample of behavior and hence, the scores or ratings derived from tests are subject to error. This is true of standardized achievement tests as well as teacher judgments, grades, observations, exams and any other measurement technique one cares to consider. Second, a

REPORTS AVAILABLE

Back issues of *Measurement in Education* are available for $1.50 each per single issue, and for 75¢ each in quantities of 25 or more for a single issue.

Vol. 1, No. 1	*Helping Teachers Use Tests* by Robert L. Thorndike
No. 2	*Interpreting Achievement Profiles — Uses and Warnings* by Eric F. Gardner
No. 3	*Mastery Learning and Mastery Testing* by Samuel T. Mayo
No. 4	*On Reporting Test Results to Community Groups* by Alden W. Badal & Edwin P. Larsen
Vol. 2, No. 1	*National Assessment Says* by Frank B. Womer
No. 2	*The PLAN System for Individualizing Education* by John C. Flanagan
No. 3	*Measurement Aspects of Performance Contracting* by Richard E. Schutz
No. 4	*The History of Grading Practices* by Louise Witmer Cureton
Vol. 3, No. 1	*Using Your Achievement Test Score Reports* by Edwin Gary Joselyn & Jack C. Merwin
No. 2	*An Item Analysis Service for Teachers* by Willard G. Warrington
No. 3	*On the Reliability of Ratings of Essay Examinations* by William E. Coffman
No. 4	*Criterion-Referenced Testing in the Classroom* by Peter W. Airasian and George F. Madaus
Vol. 4, No. 1	*Goals and Objectives in Planning and Evaluation: A Second Generation* by Victor W. Doherty and Walter E. Hathaway
No. 2	*Career Maturity* by John O. Crites
No. 3	*Assessing Educational Achievement in the Affective Domain* by Ralph W. Tyler
No. 4	*The National Test-Equating Study in Reading (The Anchor Test Study)* by Richard M. Jaeger
Vol. 5, No. 1	*The Tangled Web* by Fred F. Harcleroad
No. 2	*A Moratorium? What Kind?* by William E. Coffman
No. 3	*Evaluators, Educators, and the Publics: A Detente?* by William A. Mehrens
No. 4	*Shall We Get Rid of Grades?* by Robert L. Ebel
Vol. 6, No. 1	*On Evaluating a Project: Some Practical Suggestions* by John W. Wick
No. 2	*Measuring Reading Achievement: A Case for Criterion — Referenced Testing and Accountability* by S. Jay Samuels and Glenace E. Edwall
No. 3	*Security in a Citywide Testing Program* by Anthony J. Polemeni
No. 4	*Dear Mama: Why don't they love me anymore?* by Thomas J. Fitzgibbon
Vol. 7, No. 1	*The Discrepancy Evaluation Model (Part I)* by Andres Steinmetz
No. 2	*The Discrepancy Evaluation Model (Part II)* by Andres Steinmetz
No. 3	*What Graduate and Professional School Students Think About Admission Tests* by Leonard L. Baird
No. 4	*The Paradox of Education Testing* by Robert L. Ebel
Vol. 8, No. 1	*FACE UP: A Framework for Assessing Career Educators Utilization Programs* by Bruce W. Tuckman
No. 2	*Grade Equivalent Scores* by Gary Echternacht and *If Not Grade Equivalent Scores — Then What?* by Jeanne M. Plas
No. 3	*Criticisms of Testing: How Mean is the Median?* by Henry S. Dyer
No. 4	*The Superintendent and Testing: Implications for the Curriculum* by Herbert C. Rudman
Vol. 9, No. 1	*A Consumer's Guide to Criterion — Referenced Test Item Statistics* by Ronald C. Berk
No. 2	*Perspective on Intelligence Testing*, by Roger T. Lennon
No. 3	*Bias*, by Eric F. Gardner
No. 4	*The Use of Tests in Admissions to Higher Education* by Mary Fruen
Vol. 10,No. 1	*Some Comments on Criterior-Referenced and Norm-Referenced Achievement Tests* by William A. Mehrens and Robert L. Ebel
No. 2	*Does the Rasch Model Really Work? A Discussion for Practitioners* by R. Robert Rentz and Charlotte C. Rentz

large part of the debate over test use and misuse is best summarized as a situation in which both sides of the issue endeavor to apply the powers of logic and persuasion to support previously made emotional conclusions. The debate has reached a point where unsubstantiated generalizations about the benefits or liabilities are accepted largely uncritically. Third, there is very little empirical evidence which suggests that tests have certain consequences, either positive or negative. In the complex social environment of the classroom it is difficult if not impossible to single out one variable as the cause of other variables. Fourth, there are clear misuses of standardized achievement tests. Test results are misused when they are treated as an exact index of pupil achievement, when scores are used as the sole basis in making important decisions about individuals and programs, when users confuse the constructs that tests measure, and finally, when a single test score is interpreted in terms of factors which caused that score. Fifth, the uses of tests are somewhat fewer. Within the classroom they can provide an additional source of evidence to teachers, to be used in conjunction with other sources, to help teachers know and make decisions about pupils. Outside the classroom uses of standardized achievement tests for accountability type decisions are more problematical. The demand for various forms of accountability at the district, state, or national level, has, in most cases, outstripped the technology to satisfy these demands immediately. There is an understandable and valid demand for accountability, but there are problems in relying heavily upon existing standardized achievement tests as the primary basis for these decisions. Sixth, and last, because of all that has been said and much of what has not been said, there will be a tendency to succumb to the premise that "new or different is better" as regards alternative achievement testing strategies. This will rarely be the case, and new and different strategies of achievement testing should be judged in terms of the same concerns expressed for standardized achievement tests in this presentation. In fact, virtually this same presentation could have been made about most of the prevalent forms of achievement testing used in education.

The issue of standardized test use and misuse is not clear or straightforward. There is much we do not know about the effects of such tests on pupils, teachers, and schools. Moreover, in some of the more prevalent applications of such tests, the identification of test use or test misuse is not absolute, but contextually and perceptually influenced. If the preceding discussion has not convinced the reader of these facts, it is to be hoped at least that the reader will carry away a healthy cynicism for the over-generalized and over-simplified rhetoric of many of the proponents and opponents of testing. The truth, as was noted at the start of this discussion, lies somewhere in between.

References

Airasian, P. W. The effects of standardized testing and test information on teachers' perceptions and practices. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, April, 1979.

Airasian, P. W., Kellaghan, T., Madaus, G. F., and Pedulla, J. J. Proportion and direction of teacher rating changes of pupils' progress attributable to standardized test information. *Journal of Educational Psychology*, 1977, 69, 6, 702-709.

Airasian, P. W., Madaus, G. F., and Pedulla, J. J. *Minimal Competency Testing*. Englewood Cliffs, New Jersey: Educational Technology Publications, 1979.

Beck, M. D. and Stetz, F. P. Teachers' opinions of standardized test use and usefulness. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, April, 1979.

Bloom, B. S. Some theoretical issues relating to educational evaluation. In R. Tyler (Ed.), *Educational evaluation: New roles, new means*. The sixty-eighth yearbook for the National Society for the Study of Education, Part II. Chicago, Illinois: National Society for the Study of Education, 1969, pp. 26-50.

Brophy, J. E. and Good, T. L. *Teacher-student relationships, causes and consequences*. New York: Holt, Rinehart and Winston, 1974.

Goslin, D. A. *Teachers and testing*. New York: Russell Sage Foundation, 1967.

Madaus, G. F. and Macnamara, J. *Public Examinations*. Dublin, Ireland: The Educational Research Centre, 1970.

Madaus, G. F., Airasian, P. W., and Kellaghan, T. *School effectiveness: A reassessment of the evidence*. New York: McGraw Hill, 1980.

Rivlin, A. M. Forensic social science. *Harvard Educational Review*, 1973, 43, 1, 61-75.

Rosenthal, R. and Jacobson, L. *Pygmalion in the classroom*. New York: Holt, Rinehart and Winston, 1968.

Srinivasan, J. T. Annual terminal examinations in the Jesuit high schools of Madras, India. Unpublished doctoral dissertation, Boston College, 1971.

Stetz, F. P. and Beck, M. D. A survey of opinions concerning users of standardized tests. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto, Ontario, Canada, March, 1978.

NCME NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

Second class postage paid at East Lansing, Mich. 48823 and additional mailing points.

USPS 823120
1230 17th Street, N.W.
Washington, DC 20036